

Pneumonia binary classification using multi-scale feature classification network on chest x-ray images

Thulfiqar H. Mandeel¹, Salah M. Awad¹, Shama Naji²

¹Department of Computer techniques engineering, College of Information Technology, Imam Ja'afar Al-Sadiq University, Al-Muthanna, Iraq

²Department of Radiology, Al-Hussain Teaching Hospital, Al Muthanna, Iraq

Article Info

Article history:

Received Dec 15, 2021

Revised Apr 23, 2022

Accepted May 22, 2022

Keywords:

Convolutional neural networks
pneumonia detection

Deep learning

Transfer learning

X-ray imaging

ABSTRACT

According to the world health organization, pneumonia was the cause for 14% of all deaths of children under 5 years old. A computer-aided diagnosis (CADx) system can help the radiologist in the detection of pneumonia in chest radiographs by serving as a second opinion. The typical CADx is based on transfer learning which is done by transferring the learning of feature extraction from one task with plenty of available data to a related task with a scarcity of data. This approach has two limitations which are first, blocking the transferred model from extracting the features that are singular to the new dataset as well as the inability to reduce the complexity of the original model. To address these drawbacks, we proposed a convolutional neural network (CNN) model with low complexity and three paths for feature extraction. The proposed model extracts three different types of features and concatenates them into one feature that provides a good representation for the classes. The proposed model was evaluated on a publicly available dataset. The results showed outperformance by the proposed model compared to the transfer learning models with recall 0.912 ± 0.039 , precision 0.942 ± 0.029 , F-beta score 0.93, and Cohen's kappa score 0.740 ± 0.008 .

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Thulfiqar H. Mandeel

Department of Computer techniques engineering, College of Information Technology, Imam Ja'afar Al-Sadiq University

AL-Resala district, Samawah, Al-Muthanna, Iraq

Email: thulfiqar.hussein@sadiq.edu.iq

1. INTRODUCTION

Pneumonia illness is a lung infection caused by microbes such as bacteria or viruses. The pneumonia infection creates inflammation in the lung which results in breathing difficulty and sometimes death [1]. In 2019, a new type of pneumonia caused by the SARS-CoV-2 virus and which is called Covid-19 has spread throughout the world and turned into a pandemic [2]. At the early stages of the pandemic, the etiology was unknown which made it dangerous to people without medical profession as well as to the health workers alike. The spread of covid-19 has paralyzed the medical systems in most of the developed countries, due to its novelty and lack of medical procedures to tackle it, and the infection of health workers [3].

A computer-aided diagnosis system (CADx) is the utilization of the output of algorithms implemented on a computing device to assist radiologists in diagnosing an illness. The CADx output is used as a second opinion that complements that of the radiologist and not as a replacement [4]. The availability of a CADx system can help in reducing the contact time between the patients and the radiologist which reduces

the probability of health workers getting infections. Moreover, the CADx can help reduce the uncertainty when making diagnosis decisions by the radiologist as well as in medical education.

The research of convolutional neural network (CNN) architecture led to state-of-the-art models in multi-class classification of generic images. The benchmark was the classifying of the ImageNet database [4]. However, the deep neural network model requires an enormous amount of data and computational power to train such a model. Nevertheless, the main characteristic of the medical images datasets is that they are limited in size compared to ImageNet; Which led to the use of these state-of-the-art models as a pre-trained feature extractor [5]. Researchers have used the transfer learning models to classify X-ray images to identify COVID-19 infection [6]–[12] or pneumonia in general [5], [13]–[15], Cardiomegaly [16], osteoarthritis [17], Breast cancer [18], skin cancer [19], [20], tuberculosis detection [21], and disease-free chest [22]. Popular state-of-the-art models such as NASNet [23], ResNet101/152 [24], InceptionResNetV2 [25], and Xception [26] have been used as transfer learning models [12], [13], [18], [27] based on the hypothesis that the ImageNet features can be generalized. However, Kornblith *et al.* [28] find that the ImageNet features extractors do not generate well-discriminating features for the classification of a fine-grained dataset.

In this paper, we propose a CADx system for binary classification of chest X-ray images into two classes: Pneumonia and Normal. The proposed model contains a residual connection, two paths of convolutional layers, and multiple filters of different sizes which allow the model to extract rich discriminating features. The organization of the paper is: Section 2 describes the dataset used to train and evaluate the proposed technique. Section 3 describes the proposed model. Section 4 presents the performance evaluation, followed by the conclusion in Section 5.

2. MATERIAL AND METHOD

2.1. Chest X-ray dataset

The dataset used in this research contains 5856 validated chest X-ray images depicting pneumonia and normal cases which were collected by Kermany *et al.* [5] and publicly shared. These images are grouped into two groups which are the training- and the validation group. Table 1 shows the distribution of the images in this database and Figure 1 shows samples from this database.

Table 1. Number of images for each class in the dataset

Class	Subset	
	Training	Validation
Normal	1349	234
Pneumonia	3883	390
Total	5232	624

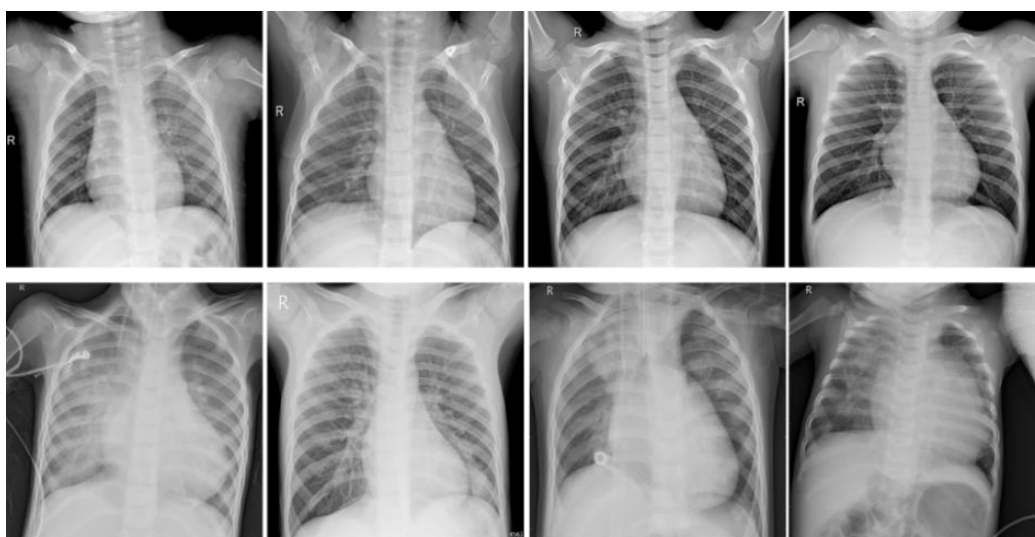


Figure 1. Sample images from the chest x-ray dataset. (the images on the first row represent normal lungs. The images on the second row represent infected lungs)

The scarcity of medical images or the scarcity of images belonging to a specific class in the dataset may affect the training process enormously by biasing the model's weight toward the class with the largest data as shown in Table 1, which results in skewed classification. Hence, a cost-sensitive learning approach is implemented in this research to deal with this problem. The cost-sensitive learning approach assigns a high cost for misclassification of the minority class while the majority class has less misclassification cost, hence, adjust the models' weights in a way that pays more attention to the minority class than to the majority class [29]. This approach didn't discard any existing images from the dataset as well as didn't generate any unreal data that didn't represent the actual pictorial information. The weights are calculated depending on the number of images for each class in the training set as in (1).

$$w_c = \frac{1}{N_c} \times \frac{N_a}{k} \quad (1)$$

where w_c is the weight for class c , N_c is the number of images in class c , N_a the total number of images in the dataset, and k is the number of the classes.

2.2. The Proposed CNN model

The proposed CNN model as shown in Figure 2, extracts three different types of features which are later concatenated into one rich feature with satisfactory classes representation. These features are extracted using three paths named A, B, C. In path A, the fine features are extracted as there is no max-pooling layer to perform down-sampling. In path B several max-pooling operations will be applied to the input of this path in sequence; In total, a sharp feature will be extracted. Moreover, path C extracts basic features which is the output from a single convolutional layer. In this model, there are four convolution layers in path A and, three in path B, and one at the beginning. In addition, a skip connection (path C) has been added to the proposed model to minimize the gradient vanishing impact. There are seven types of layers in the proposed model architecture and the details of these layers are explained in the following:

- a. The scaling layer is used to scale the pixel values of the input image in the range [-1 to 1]. After rescaling the input, a dropout rate of 0.3 is applied to expose the model to all the features in the image and prevent it from considering noise as a feature.
- b. The convolutional layer is used to apply a spatial convolutional operation on the input image with several filters which result in generating multiple feature maps. Different sizes of filters are used to capture most of the features in the original image independently from their size. A non-linear activation function is applied to the output of the convolutional operation to introduce non-linearity to capture non-linear features. In this architecture, different activation functions are used, such as Tanh, Elu, Swish, Relu, and Selu [30], [31].
- c. The batch Normalization layer is used to normalize the convolutional layers outcome to make the feature maps have unity variance and zero mean. The batch normalization layer can stabilize the learning process and reach convergence faster [32]. The batch normalization enables also reduce the initialization effect on the model weights.
- d. The Max-Pooling layer is used to downsample the image and keep the sharp features only. Three max-pooling operations with a size of 2x2 and stride of 2x2 are used in path B.
- e. The global Average Pooling layer is used to create one feature map for each class by averaging the corresponding feature maps.
- f. The concatenation layer combines the final feature map from the global average pooling in paths A, B, and C. The concatenated features will represent a singular large feature that discriminates pneumonia from the normal class.
- g. The Fully Connected layer is used to classify the concatenated features into pneumonia and normal classes depending on the threshold of the activation function. Moreover, there is a dropout rate of 0.3 is applied to the fully connected layer. The dropout introduces sparsity to the activation of the hidden neurons, i.e., sparse representation of the data and preventing the neural network from overfitting to the training data. There is no image augmentation layer in the model as it's tested in this research that it's negatively affected the performance of the models. Moreover, the use of a combination of image augmentation techniques is previously reported to increase the overfitting in very limited datasets [33].

3. RESULTS AND DISCUSSION

The metrics used in the experimental analysis are divided into metrics evaluating the performance of the model such as recall, precision, f1-score, Kappa statistics, and heat map and metrics used to evaluate the complexity of the model such as the total number of parameters, model size, and the testing time. The training and testing of the models were conducted on windows 10 using NVIDIA GeForce® GTX 1660 Ti

GPU, intel core i7-10750 2.60 GHz CPU, and 16GB of installed ram. Due to the random weight initialization in the artificial neural network, the final result usually differs each time the model is trained. Hence, we trained and tested each of the transfer-learning CNN models as well as the proposed model five times and the resulting average is calculated. The number of training epochs is 6, the batch size is 8, and the images are down-sampled to 160×160 before being feed into the models.

When designing the model, we explored a wide range of combinations of hyper-parameters against the evaluation accuracy by using Bayesian optimization with the Gaussian process. The Bayesian optimization has been used due to its performance compared to other algorithms such as grid search, random search [34]. The tuned hyper-parameters include activation functions, filters numbers and their sizes in each convolutional layer, learning rate, optimization function choice. The final hyper-parameters after tuning for the activation functions and the number of filters as well as the filter size are shown in Figure 2. Moreover, the learning rate is tuned to 0.001 and adaptive moment estimation (Adam) was chosen as the optimization function. The transfer learning models are used as a feature extractor while the classifier consists of a global average pooling layer, batch normalization, and a fully connected layer with dropout regularization.

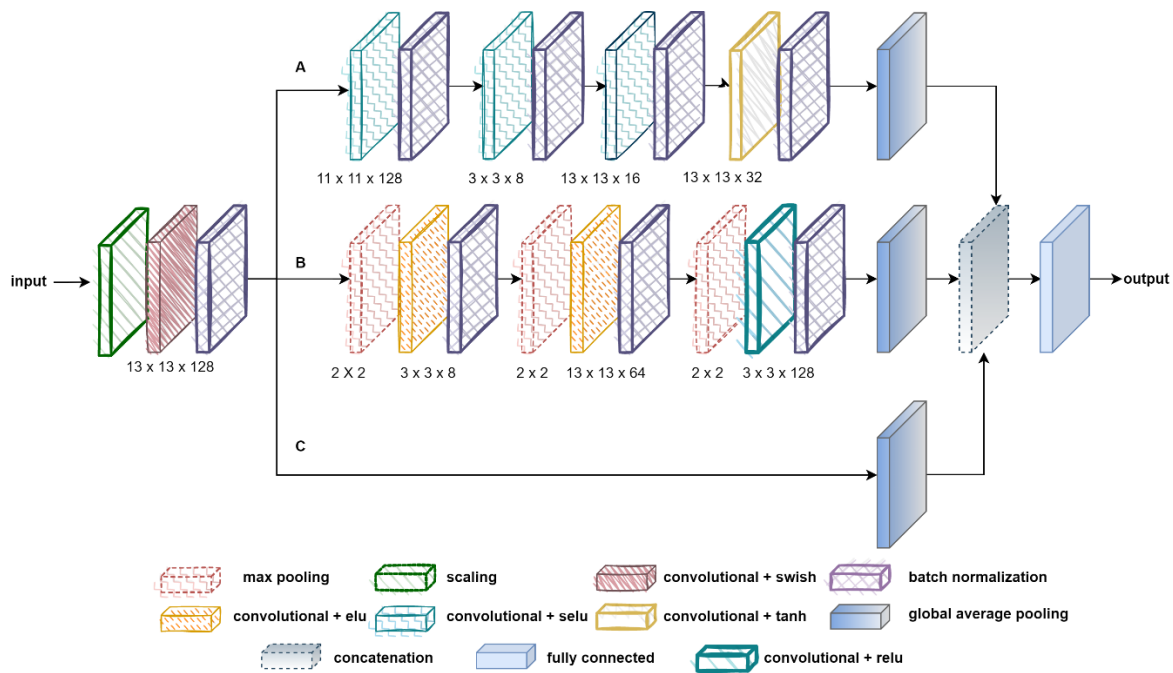


Figure 2. The proposed model

3.1. Recall, precision, and F-beta score

In the ideal situation, the input data when fed into a trained classification model, the output is the actual class of the input. In this scenario, the prediction of X-ray image of pneumonia infected class is called true positive (TP) while the prediction for a normal case is called the true negative (TN). However, in the real-case scenarios, this outcome isn't guaranteed for several reasons such as noisy data or weak handling of the features by the model, and this can produce a false positive (FP) and false negative (FN). The FN, which is a pneumonia case classified as normal, has a severe impact on the patient's health because that it will result in preventing or delaying the treatment. On the other side, FP, is a normal case classified as pneumonia, raises the medical cost and workforce. The relation between the TP and FP is called the precision and can be calculated using (2), high precision means low FP.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

while the relation between the TP and FN is called the recall which can be calculated by (3), high recall means low FN.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

The precision and recall can be used to measure the performance of the model when either the cost of FP or FN is critical consequently. In the case of diagnosis of an illness, the higher the recall is the better as the consequences are more damaging to the patient. Practically, it's hard to get a model that has both high precision and recall at the same time. The F-beta score combines the precision and recall in a weighted harmonic mean as shown in (4).

$$f - \text{Beta} = (1 + \beta^2) \cdot \frac{\text{Precision} \times \text{Recall}}{(\beta^2 \times \text{Precision}) + \text{Recall}} \quad (4)$$

The weight is chosen as $\beta = 1$ when both the precision and the recall are important. However, if we want to favor precision over the recall, then the $\beta < 1$ and otherwise $\beta > 1$. As aforementioned, the recall should be more important than the precision for the pneumonia class. Table 2 shows the performance of the proposed model along with the transfer learning models. As mentioned earlier, each test is conducted five times and the average is calculated as the final score along with the standard deviation (std).

A successful model in the case of classifying pneumonia images should have a high recall and relatively high precision. From Table 2, we can see that the Resnet101 showed the highest recall score and a low precision score for the pneumonia class. Hence, it's deemed an impractical model. While the proposed model has a combination of high scores for both the recall and precision for the pneumonia class. Which translated to a better F-beta score for all the values of β compared to the other models. The results highlight the importance of training the feature extractor on the fine-grained dataset as the chest X-ray images. The proposed model has learned a discerning feature that discriminates normal from infected lungs images which didn't exist in a generic dataset like ImageNet.

Table 2. Precision, recall, and F-beta score results

Model	Class	Precision	Recall	f - Beta		
				$\beta = 0.5$	$\beta = 1$	$\beta = 2$
InceptionResNetV2	Normal	0.712±0.025	0.864±0.009	0.74	0.78	0.83
	Pneumonia	0.904±0.004	0.79±0.004	0.88	0.84	0.81
NASNetLarge	Normal	0.844±0.042	0.784±0.095	0.83	0.81	0.80
	Pneumonia	0.878±0.045	0.908±0.041	0.88	0.89	0.90
ResNet101V2	Normal	0.874±0.021	0.702±0.058	0.83	0.78	0.73
	Pneumonia	0.844±0.024	0.938±0.012	0.86	0.89	0.92
ResNet152V2	Normal	0.868±0.039	0.804±0.077	0.85	0.83	0.82
	Pneumonia	0.89±0.038	0.924±0.036	0.90	0.91	0.92
Xception	Normal	0.838±0.031	0.82±0.036	0.83	0.83	0.82
	Pneumonia	0.894±0.018	0.902±0.025	0.90	0.90	0.90
Proposed model	Normal	0.866±0.050	0.902±0.056	0.87	0.88	0.89
	Pneumonia	0.942±0.029	0.912±0.039	0.94	0.93	0.92
Radiologist	Normal	0.96	0.91	0.95	0.93	0.92
	Pneumonia	0.95	0.98	0.96	0.96	0.97

3.2. Kappa statistics

Cohen's kappa statistic (interrater reliability) is a measurement to test the agreement between different raters given the same data [35]. Despite using the same data, the used CNN models in this study have different feature extraction methods which result in a different classification. Hence, the kappa statistic is used to determine the interrater reliability between the radiologist and the CNN models. The kappa statistic can be calculated as in (5). The range for the kappa statistic is from -1 to +1. The interpretations of Kappa score are: when the score ≤ 0 then there is no agreement between the raters; when the score is between 0.01 and 0.20 then there is a slight agreement. If the score is between 0.21 and 0.40, then there is a fair agreement; the score between 0.41 and 0.60 shows a moderate agreement. A score between 0.61 and 0.80 means substantial agreement; finally, a score between 0.81 to 1.00 means almost perfect agreement. Table 3 shows Cohen's kappa statistic for the CNN models.

$$\kappa = \frac{Pr(a) - Pr(e)}{Pr(e)} \quad (5)$$

where $Pr(a)$ is the probability of agreement, and $Pr(e)$ is the expected agreement.

Table 3. Kappa statistic

Model	Kappa statistics
InceptionResNetV2	0.583±0.030
NASNetLarge	0.648±0.042
ResNet101V2	0.614±0.029
ResNet152V2	0.688±0.050
Xception	0.676±0.033
Proposed model	0.740±0.008

From Table 3, we can see that most of the models have a substantial agreement between the radiologist assessment and the models' predictions. However, the proposed model has the highest kappa score with the least standard deviations. The kappa statistic result might enforce the idea that the proposed model has learned medical diagnostic features.

3.3. Class activation map (CAM)

The class activation map (CAM) is a helpful tool to find what features in the image that has the highest impact on the prediction of image class. Hence, the CAM can be used to validate that the model is picking the right underlying pattern for each class. If the model didn't pick the right patterns, the training of the model should be revised. However, CAM has a drawback which is the requirement for changing the CNN model and dropping the fully connected layers. Selvaraju *et al.* [36] proposed a generalization to CAM called the gradient-weighted class activation map (Grad-CAM). The Grad-CAM didn't require architecture modification as well as it is applicable for a wide range of CNN-models families. The Grad-CAM uses the gradient information that feeds into the last convolutional layer of the CNN to visualize the importance of the image's parts for the classification of the class at hand.

Figures 3 and 4 show the X-rays images and the corresponding Grad-CAM. Note that in the normal cases on X-ray images, the lungs appear with a dark shade, whereas the spine appears with white shade; This is because the air in the lungs has smaller attenuation compared with bones in the spine. Generally, radiologists diagnose pneumonia when there is a loss-of-silhouette sign [37] which is the loss of the heart borders with the adjacent lungs segments. It is observed that the normal and pneumonia Grad-CAMs can highlight the medial part of the X-ray images, including lungs and part of the spine. However, The pneumonia Grad-CAMs show a significant loss-of-silhouette sign in X-ray images.

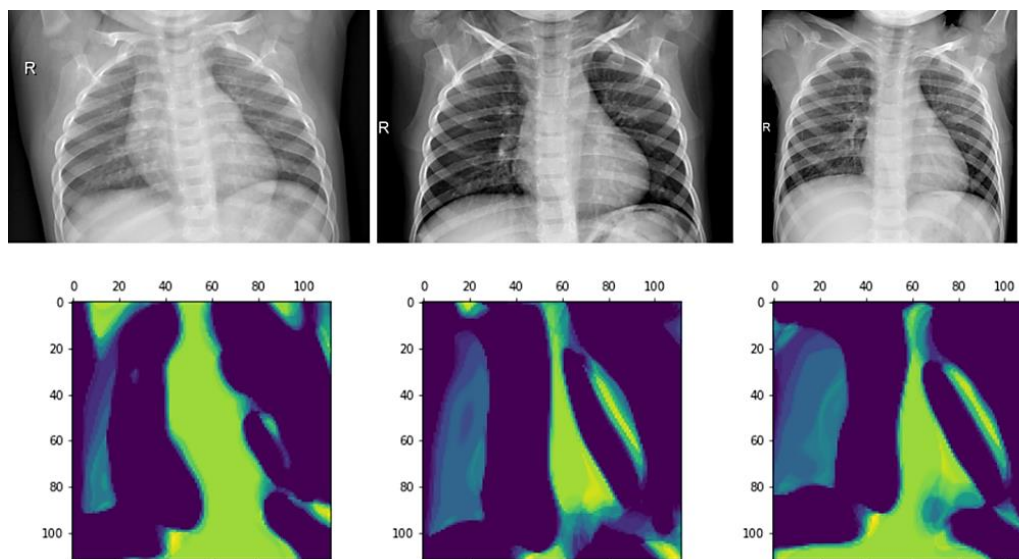


Figure 3. Normal images. first row: chest x-ray images, second row: Corresponding heat map

3.4. Model complexity

A model with a low level of complexity is more suitable to be deployed in real-time CADx as well as deployment on devices with low computational capabilities and storage size. The main factor that affects the model complexity is the number of model parameters. Model parameters are the variables that are tuned

throughout the training of the artificial neural network training to enhance the prediction of the model. The higher the number of the parameter the more computations are required in both training and testing. The time required for training and testing is the by-product of the number of parameters that the model has. The proposed model has a very small size compared to the transfer learning models as can be seen in Table 4. This size will also affect the time of testing chest X-ray images, as the number of addition-multiplication operations will decrease. The test time in Table 4 is done for a batch of size 32.

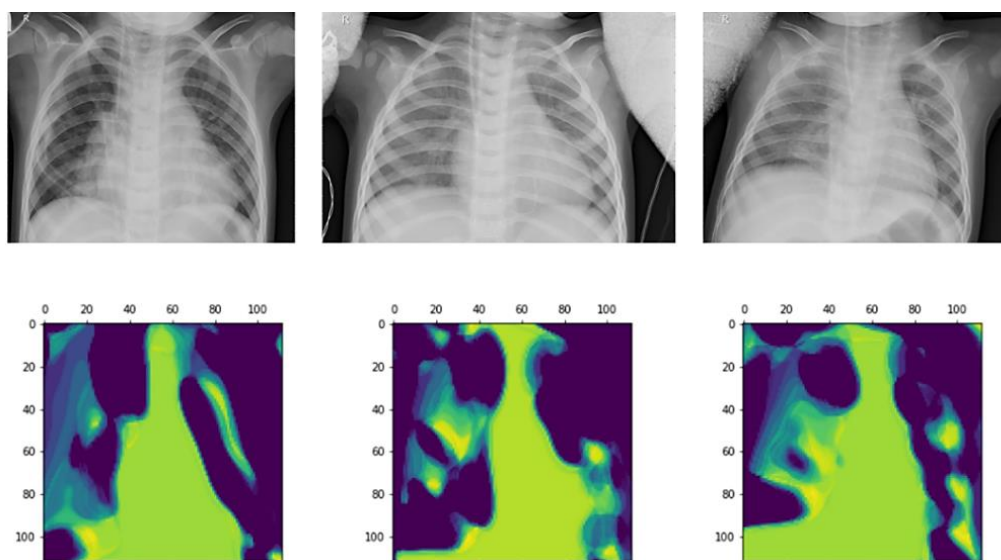


Figure 4. Pneumonia images. first row: Chest x-ray images, second row: corresponding heat map

Table 4. Models' complexity comparison

Model	Total # of parameters	Size (in KB)	Time (in Sec)
InceptionResNetV2	54,344,417	214,377	2.99±0.05
NASNetLarge	84,936,979	334,750	4.44±0.11
ResNet101V2	42,636,801	167,619	1.79±0.08
ResNet152V2	58,341,889	229,467	2.41±0.06
Xception	20,871,721	81,992	1.19±0.06
Proposed model	2,337,057	27,548	0.87±0.03

4. CONCLUSION




In this paper, we proposed a new CNN model which tested on pneumonia binary classification, and the results were verified using recall, precision, β -score, Kappa statistics. The model is more suitable for real-time CADx as it has lower complexity compared to transfer learning models. Unlike the transfer learning models which rely on generic features, the proposed model was trained to extract a fine-grained feature in the dataset. The extracted features gave the proposed model a lead in the test scores as it effectively represents the Pneumonia and normal classes. The extracted features are proved to be correct by observing the Grad-CAM that is extracted from the advanced convolutional layers in the model.

REFERENCES




- [1] D. Bamford and M. Zuckerman, "Encyclopedia of virology fourth edition," p. 4109, 2021.
- [2] "Coronavirus," 2022, [Online]. Available: https://www.who.int/health-topics/coronavirus#tab=tab_1.
- [3] W. H. O. World Health Organization, "Coronavirus (COVID-19) events as they happen," WHO, 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>.
- [4] L. Beyer, O. J. Hénaff, A. Kolesnikov, X. Zhai, and A. van den Oord, "Are we done with ImageNet?," 2020, [Online]. Available: <http://arxiv.org/abs/2006.07159>.
- [5] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122--1131.e9, 2018, doi: 10.1016/j.cell.2018.02.010.
- [6] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. Jamalipour Soufi, "Deep-COVID: predicting COVID-19 from chest X-ray images using deep transfer learning," *Medical Image Analysis*, vol. 65, 2020, doi: 10.1016/j.media.2020.101794.
- [7] B. Narayanan, R. Hardie, V. Krishnaraja, C. Karam, and V. Davuluru, "Transfer-to-transfer learning approach for computer aided detection of COVID-19 in chest radiographs," *Ai*, vol. 1, no. 4, pp. 539--557, 2020, doi: 10.3390/ai1040032.

- [8] A. Shamsi *et al.*, “An uncertainty-aware transfer learning-based framework for covid-19 diagnosis,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 4, pp. 1408–1417, 2021, doi: 10.1109/TNNLS.2021.3054306.
- [9] M. M. Rahaman *et al.*, “Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning approaches,” *Journal of X-Ray Science and Technology*, vol. 28, no. 5, pp. 821–839, 2020, doi: 10.3233/XST-200715.
- [10] H. Benbrahim, H. Hachimi, and A. Amine, “Deep transfer learning pipelines with apache spark and keras tensorflow combined with logistic regression to detect covid-19 in chest ct images,” *Walailak Journal of Science and Technology*, vol. 18, no. 11, 2021, doi: 10.48048/WJST.2021.13109.
- [11] N. Narayan Das, N. Kumar, M. Kaur, V. Kumar, and D. Singh, “Automated Deep Transfer Learning-Based Approach for Detection of COVID-19 Infection in Chest X-rays,” *Irbm*, vol. 43, no. 2, pp. 114–119, 2022, doi: 10.1016/j.irbm.2020.07.001.
- [12] I. D. Apostolopoulos and T. A. Mpesiana, “Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks,” *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, pp. 635–640, Jun. 2020, doi: 10.1007/s13246-020-00865-4.
- [13] J. E. Luján-García, C. Yáñez-Márquez, Y. Villuendas-Rey, and O. Camacho-Nieto, “A transfer learning method for pneumonia classification and visualization,” *Applied Sciences (Switzerland)*, vol. 10, no. 8, 2020, doi: 10.3390/APP10082908.
- [14] V. Chouhan *et al.*, “A novel transfer learning based approach for pneumonia detection in chest X-ray images,” *Applied Sciences (Switzerland)*, vol. 10, no. 2, 2020, doi: 10.3390/app10020559.
- [15] T. Rahman *et al.*, “Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray,” *Applied Sciences (Switzerland)*, vol. 10, no. 9, 2020, doi: 10.3390/app10093233.
- [16] S. Zhou, X. Zhang, and R. Zhang, “Identifying cardiomegaly in ChestX-ray8 using transfer learning,” *Studies in Health Technology and Informatics*, vol. 264, pp. 482–486, 2019, doi: 10.3233/SHTI190268.
- [17] P. S. Q. Yeoh *et al.*, “Emergence of deep learning in knee osteoarthritis diagnosis,” *Computational Intelligence and Neuroscience*, vol. 2021, 2021, doi: 10.1155/2021/4931437.
- [18] L. G. Falconi, M. Perez, and W. G. Aguilar, “Transfer learning in breast mammogram abnormalities classification with Mobilenet and Nasnet,” *International Conference on Systems, Signals, and Image Processing*, vol. 2019-June, pp. 109–114, 2019, doi: 10.1109/IWSSIP.2019.8787295.
- [19] S. Alheejawi, R. Berendt, N. Jha, S. P. Maity, and M. Mandal, “Detection of malignant melanoma in H&E-stained images using deep learning techniques,” *Tissue and Cell*, vol. 73, 2021, doi: 10.1016/j.tice.2021.101659.
- [20] R. A. Pratiwi, S. Nurmaini, D. P. Rini, M. N. Rachmatullah, and A. Darmawahyuni, “Deep ensemble learning for skin lesions classification with convolutional neural network,” *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 3, pp. 563–570, Sep. 2021, doi: 10.11591/ijai.v10.i3.pp563-570.
- [21] S. K. T. Hwa, M. H. A. Hijazi, A. Bade, R. Yaakob, and M. S. Jeffree, “Ensemble deep learning for tuberculosis detection using chest X-ray and canny edge detected images,” *IAES International Journal of Artificial Intelligence*, vol. 8, no. 4, pp. 429–435, 2019, doi: 10.11591/ijai.v8.i4.pp429-435.
- [22] M. Moradi, K. C. L. Wong, T. Syeda-Mahmood, and J. T. Wu, “Identifying disease-free chest x-ray images with deep transfer learning,” p. 24, 2019, doi: 10.1117/12.2513164.
- [23] B. Zoph, V. Vasudevan, J. Shlens, and Q. V Le, “Learning transferable architectures for scalable image recognition,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 8697–8710, 2018, doi: 10.1109/CVPR.2018.00907.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9908 LNCS, pp. 630–645, 2016, doi: 10.1007/978-3-319-46493-0_38.
- [25] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, inception-ResNet and the impact of residual connections on learning,” *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, vol. 31, pp. 4278–4284, 2017.
- [26] F. Chollet, “Xception: deep learning with depthwise separable convolutions,” *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 1800–1807, 2017, doi: 10.1109/CVPR.2017.195.
- [27] M. Saini and S. Susan, “Data augmentation of minority class with transfer learning for classification of imbalanced breast cancer dataset using inception-V3,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11867 LNCS, pp. 409–420, 2019, doi: 10.1007/978-3-030-31332-6_36.
- [28] S. Kornblith, J. Shlens, and Q. V. Le, “Do better ImageNet models transfer better?,” *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2019-June, pp. 2656–2666, Jun. 2009, doi: 10.48550/arXiv.1805.08974.
- [29] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, “Learning from class-imbalanced data: Review of methods and applications,” *Expert Systems with Applications*, vol. 73, pp. 220–239, 2017, doi: 10.1016/j.eswa.2016.12.035.
- [30] M. M. Lau and K. Hann Lim, “Review of adaptive activation function in deep neural network,” pp. 686–690, 2019, doi: 10.1109/iecbes.2018.8626714.
- [31] B. Zoph and Q. V Le, “Searching for activation functions,” *6th Int. Conf. Learn. Represent. ICLR 2018 - Work. Track Proc.*, vol. 6, pp. 1–13, 2018, doi: 10.48550/arXiv.1710.05941.
- [32] B. L. G. Chen, P. Chen, Y. Shi, C.-Y. C.-Y. Hsieh and S. Zhang, “Rethinking the usage of batch normalization and dropout in the training of deep neural networks,” 2019, doi: 10.48550/arXiv.1905.05928.
- [33] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.
- [34] L. K. F. Hutter and J. Vanschoren, “Automated learning machine methods, systems, challenges,” 2020, doi: 10.1007/978-3-030-05318-5.
- [35] M. M. Islam, M. A. Kashem, and J. Uddin, “Fish survival prediction in an aquatic environment using random forest model,” *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 3, pp. 614–622, Sep. 2021, doi: 10.11591/ijai.v10.i3.pp614-622.
- [36] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: visual explanations from deep networks via gradient-based localization,” *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, 2020, doi: 10.1007/s11263-019-01228-7.
- [37] T. Bansal and R. Beese, “Interpreting a chest X-ray,” *British Journal of Hospital Medicine*, vol. 80, no. 5, pp. C75--C79, 2019, doi: 10.12968/hmed.2019.80.5.c75.




BIOGRAPHIES OF AUTHORS

Thulfiqar H. Mandeel    received the M.Sc. degree in Embedded system design engineering and Ph.D. degree in computer engineering from Universiti Malaysia Perlis (UniMAP) in Perlis, Malaysia in 2015 and 2019 respectively. He joined UniMAP in 2016 as a graduate assistant, and is currently a lecturer in Imam Ja'afar Al Sadiq University in Iraq. His current research interests include image processing, biometric recognition, medical image analysis, computer vision and deep learning. He can be contacted at email: Thulfiqar.hussein@sadiq.edu.iq.



Salah M. Awad    received his M-Tech degree in Communication & Information System from Aligarh Muslim University, India, in 2013 and his PhD degree in Digital Signal and Image Processing from electrical and computer engineering department, University of Alberta, Edmonton, AB, Canada, in 2021. Currently, he is a lecturer at Imam Ja'afar Al Sadiq University in Iraq and his research interests include medical image analysis, pattern recognition, and computer vision. He can be contacted at email: Email: salah.m11@sadiq.edu.iq.



Shama Naji    Graduated from medical college of Baghdad in 2002. She attained radiology diploma in 2006. She joined radiology arab board study between 2012-2016. She currently works as as a senior radiologist at Al Husain teaching hospital. She can be contacted at email: shama.naji@gmail.com.